

## **SYSTEM AND METHOD FOR MEASURING AND RECORDING LATENCY IN INTERNET PROTOCOL NETWORKS**

### **FIELD OF THE INVENTION**

**[0001]** The present invention relates generally to data networks, such as internet protocol networks, and more particularly, to a system and method for measuring and recording latency in a data network.

### **BACKGROUND OF THE INVENTION**

**[0002]** While tools do exist today which measure latency in a data network, such as a wireless data network, this functionality is not integrated into the set of services provided by wireless networks today. Furthermore these tools are used to measure latency at the equipment level, not at the user session level.

**[0003]** In wireless networks, a mobile station is connected to a radio access node via a wireless or radio frequency (RF) network. The radio access node is then connected to the data network (e.g., the Internet) typically via a wire-line network. The radio frequency network is frequently the most unreliable part of the end-to-end connection. Bad connections or a high volume of voice calls in the sector will affect the quality of data calls and thus increase latency. Thus, it is important for service providers to be able to measure wireless access latency at the user session level from the mobile station to the radio access node (i.e., the packet data serving node or PDSN).

**[0004]** Accounting information is generated for data calls, providing information such as the number of bytes transferred for a given data session and the session duration. However, currently nothing is generated which measures user experience on a per user, per session basis.

[0005] Accordingly, there is a need for a system and method for measuring and recording latency in a data network on a per user, per session basis.

## **SUMMARY OF THE INVENTION**

[0006] A system and method for measuring and recording latency in data networks, such as internet protocol (IP) networks. The system and method can be used to measure and record latency in a Simple IP network or a Mobile IP network. One aspect of the invention utilizes control plane messages for measuring latency to avoid the messages associated with the latency measurements from being billed to the user.

[0007] One aspect of the invention is a system and method for measuring and recording latency between a mobile station (MS) and a packet data serving node (PDSN) or foreign agent (FA). Another aspect of the invention is for measuring and recording latency between the PDSN/FA and a home agent (HA). Latency information can be made available for billing, statistical and reporting purposes.

[0008] One embodiment of the present invention is a system and method for measuring latency between a first device and a second device, the first and second devices communicating in accordance with a communications specification. The method comprises the second device transmitting, during a communication session, a message to the first device; receiving a response message from the first device; computing an elapsed time from transmission of the message to receipt of the response message to determine the latency; and recording the latency in a latency parameter in a data record. The system comprises a first device and a second device adapted for communicating in accordance with a communications specification. The second device is adapted for transmitting, during a communication session, a message to the first device; receiving a response message from the first device; computing an elapsed time from transmission of the message to receipt of the response message to determine the latency; and recording the latency in a latency parameter in a data record. The message and the response message are provided by the communications specification.

**[0009]** Another embodiment of the present invention is a system and method, performed by a packet data serving node, for measuring latency. The method comprises the packet data serving node storing a first start time; transmitting, to a mobile station, a Link Control Protocol Echo message; receiving a Link Control Protocol Echo Response message from the mobile station; storing a first stop time; and computing a wireless access latency based on the first start time and the first stop time. In a related embodiment, the method further comprises the packet data serving node storing a second start time; transmitting, to a home agent, a Mobile IP Registration Request message; receiving a Mobile IP Registration Reply message from the home agent; storing a second stop time; and computing an internet access latency based on the second start time and the second stop time. In another related embodiment, the step of computing the internet access latency further comprises adjusting the internet access latency for a processing time associated with the home agent.

**[0010]** The system comprises a mobile station and a packet data serving node. The packet data serving node is adapted for wirelessly communicating with the mobile station, transmitting a link control protocol echo message to the mobile station, receiving a link control protocol response message from the mobile station, and computing an elapsed time from transmission of the link control protocol echo message to receipt of the link control protocol response message to determine the wireless access latency. In a related embodiment, the system comprises a home agent and a packet data serving node for communicating with the home agent. The packet data serving node is adapted for transmitting a mobile internet protocol registration request message to the home agent, receiving a mobile internet protocol registration reply message from the home agent, and computing an elapsed time from transmission of the mobile internet protocol registration request message to receipt of the mobile internet protocol registration reply message to determine the internet access latency.

## **BRIEF DESCRIPTION OF THE DRAWINGS**

**[0011]** A more complete understanding of the present invention and its advantages will be readily apparent from the following Detailed Description taken in conjunction with the

accompanying drawings. Throughout the accompanying drawings, like parts are designated by like reference numbers and in which:

Fig. 1 is a schematic illustration of a data network in accordance with the present invention;

Fig. 2 is a flow diagram illustrating a message sequence for measuring wireless access latency in a Simple IP network in accordance with the present invention; and

Figs. 3A and 3B are flow diagrams illustrating a message sequence for measuring internet access latency and wireless access latency in a Mobile IP network in accordance with the present invention.

## **DETAILED DESCRIPTION**

**[0012]** Fig. 1 schematically illustrates a representative environment of the present invention. A mobile station (MS) 100 (e.g., a cellular phone or a wireless data terminal) is wirelessly connected to a base station (BS) 105 via a radio frequency (RF) network. The base station 105 is connected to a base station controller (BSC) 110, typically via a wire-line, such as a T1 line. The base station controller 110 is configured for connecting to and controlling multiple base stations 105. The base station controller 110 is connected to a Packet Control Function (PCF) 115, which typically resides proximate the base station controller 110. The base station controller 110 is also connected to a Mobile Switching Center (MSC) 120, which is connected to a Home Location Register (HLR) 125.

**[0013]** The PCF 115 is connected to a mobility agent, such as a packet data serving node (PDSN) 130 via the data network using the Radio Protocol (RP). The PDSN 130 is connected to an accounting authentication and authorization (AAA) server 135 via the data network using the Remote Authentication Dial In User Service (RADIUS) protocol. In a Simple IP network the PDSN 130 is connected to an Internet Protocol (IP) network, such as the Internet. In a Mobile IP system, the PDSN 130 is also connected to a Home Agent (HA) 145 via the data network 140, using Mobile IP protocol.

**[0014]** In one embodiment, the PDSN 130 provides the mobile station 100 with access to the Internet, intranets and application servers utilizing a cdma2000 Radio Access Network. The PDSN 130 is capable of performing two basic functions: (1) exchanging packets with the mobile station 100 over the RF network, and (2) exchanging packets with other IP networks 140. The PDSN 130 is capable of interfacing with the AAA server 135 to provide the mobile station 100 with a gateway to the IP network 140.

**[0015]** In Mobile IP applications, wherein the mobile station 100 is capable of roaming outside of its home network, the PDSN 130 is capable of interfacing with home agents (HA) 145. An HA 145 is typically a router located on the mobile station's home network, which is capable of tunneling packets to the mobile station while it is away from the home network. The PDSN 130 is capable of acting as a foreign agent (FA) on behalf of home agents to enable service providers to provide Mobile IP services to roaming mobile stations 100. The FA can register the mobile station 100 with their particular HA 145 and provide a forwarding address for data delivery. The PDSN/FA 130 is typically a router located on a foreign network that is capable of de-tunneling the packets from the HA 145 and providing the packets to the mobile station 100 via the RF network.

**[0016]** In data applications, a data call is typically routed through a voice network. The Mobile Switching Center 120, in conjunction with the a Home Location Register (HLR) 125 determines that the call is a data call and sends the call back to the Base Station Controller 110, which then routes the call through the PDSN 130 to the data network 140.

**[0017]** Fig. 2 illustrates a message sequence for measuring wireless access latency in a simple IP network. In one embodiment of the present invention, a radio access node or mobility agent like the PDSN can perform measurement of wireless access latency. The radio access node can compute a round trip delay between itself and the mobile station by measuring a time interval from the time the radio access node sends a message to the mobile station to the time the radio access node receives a response from the mobile station. In one embodiment, the message and response are provided by a communications standard for use between the radio access node and the mobile station. In a preferred embodiment, the

message and response are control plane messages which do not affect session data usage (billable bytes over bearer) and thus do not increase the end user charges.

**[0018]** In a preferred embodiment wherein the radio access node is a PDSN 130 utilizing cdma2000 in accordance with the TIA/IS 835-B standard, the PDSN 130 utilizes link control protocol (LCP) messages as outlined in TIA/IS 835-B to measure the wireless access latency. The round trip delay or FA to MS Latency Time in milliseconds (FMLT), is the time interval measured from the time the PDSN/FA 130 sends the mobile station 100 an LCP Echo Request message to the time the PDSN/FA 130 receives the LCP Echo Response message from the mobile station 100.

**[0019]** The TIA/IS 835-B standard provides for the LCP Echo Request and Response messages. Thus, the PDSNs 130 and mobile stations 100 support the LCP messages. The TIA/IS 835-B standard outlines using these LCP messages only in the context of detecting mobile station 100 availability following point-to-point protocol (PPP) session inactivity. The present invention, however, advantageously uses these LCP messages to measure wireless access latency. Utilizing messages provided for by the standard allows the system and method of the present invention to be compatible with existing standards, and further, to be backward compatible with existing deployed solutions. Using messages provided by the standard simplifies implementation in the currently deployed wireless network infrastructure equipment and also minimizes taxing the network infrastructure.

**[0020]** The LCP Echo messages can be sent to the mobile station 100 and used to measure wireless access latency under the following conditions: (1) when a point-to-point protocol (PPP) session is successfully established; (2) when the PPP Inactivity Timer expires; and optionally, (3) when a configurable Wireless Access Latency Timer in the PDSN expires.

**[0021]** In one embodiment, an LCP Echo Request/Response message exchange is sent when the PPP connection is established. While this message exchange is provided for by the TIA/IS 835-B standard, the present invention advantageously monitors both the time that the LCP Echo Request message is sent and the time that the LCP Echo Response message is

received in order to measure the wireless access latency. The elapsed time (in milliseconds) from the sending of the LCP Echo Request message to the receipt of the LCP Echo Response message can then be recorded in the FMLT parameter for transmission to the AAA server 135. Repeated exchanges of these messages between the mobile station 100 and the PDSN 130 can then be sent when the PPP Inactivity Timer expires, and optionally, when the Wireless Access Latency Timer expires.

**[0022]** The Wireless Access Latency Timer is not provided by the TIA/IS 835-B standard, but can be implemented in the PDSN 130 and used to trigger the wireless access latency measurement. The PPP Inactivity Timer is provided in the TIA/IS 835-B standard, and can be used in addition to or instead of the Wireless Access Latency Timer (depending on whether the Wireless Access Latency Timer is implemented) to trigger the LCP Echo Request message to the mobile station 100 after a PPP session is established. Because these timers are configurable in the PDSN 130, the service provider can have complete control over the number of samples desired to compute latency at the user session level.

**[0023]** The RADIUS protocol provides for a Usage Data Record (UDR) having certain parameters, which can be used for accounting and billing purposes. In one embodiment, start, stop, and interim usage data records are transmitted from the PDSN to the AAA server at the start of a communications session, at the end of the communications session, and during the communications session.

**[0024]** The RADIUS protocol also allows new parameters to be added to the UDR. In one embodiment, the PDSN 130 is capable of adding a new accounting parameter attribute, the FMLT (also part of this invention), as a field in the Usage Data Record (UDR) of the PDSN 130. The PDSN 130 can then send the UDR, containing the FMLT information, to the AAA server 135. The FMLT accounting parameter can be added, for example, in a Quality of Service section of the UDR. By adding the FMLT parameter to the UDR, latency information can thus be recorded in the AAA server 135 on a per-user, per-session basis. Recording of latency information can be advantageously performed in the context of an existing framework. The AAA server 135 can store the FMLT information for further

processing by other network management entities (e.g., for billing, statistical and/or reporting purposes).

**[0025]** The LCP Echo Request/Response messages are control plane messages and thus will not be included as part of the data bearer usage calculations. Using the LCP messages to measure latency will not affect a user's session data usage and the user will not be billed for these messages.

**[0026]** The process for measuring Wireless Access Latency proceeds as follows. Once the PPP setup between the mobile station 100 and the PDSN/FA 130 has been completed (step 200), the PDSN/FA 130 sends a RADIUS Accounting Request Start message to the AAA server 135 (step 205). The AAA server 135 responds with a RADIUS Accounting Response message to the PDSN/FA 130 (step 210). The PDSN/FA 130 then gets and stores the current time stamp for this session in a StartTime parameter (step 215). The PDSN/FA 130 sends the LCP Echo Request message to the mobile station 100 (step 220). The mobile station 100 receives the LCP Echo Request message and, in response, returns the LCP Echo Response message to the PDSN/FA 130.

**[0027]** The PDSN/FA 130 receives the LCP Echo Response message (step 235) and gets and stores the current time stamp for this session in a StopTime parameter (step 240). The PDSN/FA 130 then computes the Wireless Access Latency using the StartTime and StopTime parameters (step 245). The computed Wireless Access Latency is then assigned to an FMLT accounting parameter in, for example, a UDR of the PDSN 130. The PDSN 130 can then send the UDR with the FMLT parameter to the AAA server 135 in a RADIUS Accounting Request message (step 255). Once the message has been received, the AAA server 136 returns a RADIUS Accounting Response message to the PDSN 130 (step 260). No new functionality is required to be in the AAA server 135. The AAA server 135 can simply store the UDR for further processing by other network management entities.

**[0028]** Figs. 3A and 3B illustrate a message sequence for measuring latency in a mobile IP session. In mobile IP, an IP address of the mobile station 100 is anchored at its home agent



(HA) 145. The mobile station 100 is thus able to keep the same IP address, and its connectivity to the network, as it moves from one PDSN 130 to another. The mobile station's HA 145 forwards messages destined for the mobile station 100 to a foreign agent (FA) 130 via a Mobile IP network. The FA 130 then routes the messages to the mobile station 100 using the RF network.

**[0029]** Two types of latency exist in Mobile IP networks: Wireless Access Latency and Internet Access Latency. Wireless Access Latency is measured using the FMLT parameter discussed above. Internet Access Latency can be measured by utilizing messages transmitted between the FA 130 and the HA 145 (in accordance with a communications standard), taking into account any processing time required by the HA 145 before responding to the message.

**[0030]** In one embodiment, the messages used for measuring Internet Access Latency are control plane messages which do not affect session data usage (billable bytes over bearer) and thus do not increase the end user charges. In a preferred embodiment, Internet Access Latency can be measured using Mobile IP (MIP) Registration messages outlined in TIA/IS 835-B. Utilizing messages provided for by the standard allows the system and method of the present invention to be compatible with existing standards, and further, to be backward compatible with existing deployed solutions. Using messages provided by the standard simplifies implementation in the currently deployed wireless network infrastructure equipment and also minimizes taxing the network infrastructure.

**[0031]** In the context of Internet Access Latency, the round trip delay or FA to HA Latency Time in milliseconds (FHLT), is the time interval measured from the time the PDSN (acting as an FA) 130 sends the HA 145 a Mobile IP (MIP) Registration Request Message to the time the PDSN/FA 130 receives the MIP Registration Reply message from the HA 145. The HA 145 may require some processing time to process the MIP Registration Request Message, before returning the MIP Registration Reply message. If so, the processing time required by the HA 145 (or an estimate of the processing time) can be subtracted from the total round trip delay to arrive at the FHLT. The MIP Registration Request/Reply messages are provided in TIA/IS 835-B for establishing and maintaining a Mobile IP user session between the mobile

station 100 and the HA 145. The present invention, however, advantageously uses these MIP Registration messages to measure Internet access latency up to the edge of the service provider network where the HA node is located. In general the MS, based on TIA/IS 835-B, will register with the HA:

After receiving an Agent Advertisement from the PDSN due to initial establishment of a PPP session with this PDSN; and

Before expiration of the MIP Lifetime timer in the PDSN and HA.

**[0032]** In one aspect of the invention for a Mobile IP scenario, the LCP Echo messages are used to measure wireless access latency as was the case for Simple IP under the following conditions: (1) after reception of Mobile IP Registration Reply; (2) when the PPP Inactivity Timer expires; and optionally, (3) when a configurable Wireless Access Latency Timer in the PDSN expires.

**[0033]** The Wireless Access Latency Timer is not provided by the TIA/IS 835-B standard, but can be implemented in the PDSN 130 and used to trigger the wireless access latency measurement. The PPP Inactivity Timer is provided in TIA/IS 835-B standard, and can be used in addition to or instead of the Wireless Access Latency Timer (depending on whether the Wireless Access Latency Timer is implemented) to trigger the LCP Echo message to the MS. Because these timers are configurable in the PDSN 130, the service provider can have complete control over the number of samples desired to compute latency at the user level.

**[0034]** In one embodiment, the PDSN 130 is capable of adding a new accounting parameter attribute, the FHLT (round trip delay from FA to HA in milliseconds), to the Usage Data Record (UDR) of the PDSN 130. The PDSN 130 can then send the UDR, containing the FHLT information (and optionally the FMLT information), to the AAA server 135. The FHLT accounting parameter can be added, for example, in a Quality of Service section of the UDR. By adding the FHLT parameter to the UDR, latency information can thus be recorded in the AAA server 135 on a per-user, per-session basis and in the context of an existing framework. The AAA server 135 can store the FHLT information (and optionally the FMLT

information) for further processing by other network management entities (e.g., for billing, statistical and/or reporting purposes).

**[0035]** The MIP Registration messages are control plane messages. Thus, the messages will not be included, as part of the data bearer usage calculations and the user will not be billed for these messages.

**[0036]** The process for measuring wireless and internet access latency in a mobile IP network proceeds as follows. Once the PPP setup between the mobile station 100 and the PDSN/FA 130 has been completed (step 300), the FA 130 advertises its existence to the mobile station 100 (step 305). The mobile station 100 is thus able to discover whether it is at home or away from home depending on whether it is communicating with an HA 145 or a FA 130. The mobile station 100 then sends an MIP Registration Request to the PDSN/FA 130 (step 310). The PDSN/FA 130 then authenticates the MS using a RADIUS Access Request message to the AAA server 135 (step 315). The AAA server 135 responds with a RADIUS Access Accept message (step 320).

**[0037]** The PDSN/FA 130 gets and stores the current time stamp for this session in a StartTime parameter (step 325) and sends an MIP Registration Request to the HA 145 (step 330). The HA 145 receives the MIP Registration Request and, after processing the request (step 335), returns an MIP Registration Reply message to the PDSN/FA 130. The PDSN/FA 130 receives the MIP Registration Reply message (step 340) and gets and stores the current time stamp for this session in a StopTime parameter (step 345). The PDSN/FA 130 then computes the Internet Access Latency using the StartTime and StopTime parameters, adjusting the result by an estimate of the time required by the HA to process the request (i.e., the estimated HA processing time) (step 350). The computed Internet Access Latency is then assigned to an FHLT accounting parameter in, for example, a UDR of the PDSN/FA 130. After the PDSN/FA 130 receives the MIP Registration Reply from the HA 145, the PDSN/FA 130 sends the MIP Registration Reply to the mobile station 100 (step 355).

**[0038]** The PDSN/FA 130 then sends a RADIUS Accounting Request Start message containing the FHLT parameter to the AAA server 135 (step 360). The AAA server 1350 responds with a RADIUS Access Accept message (step 365). The PDSN/FA 130 gets and stores the current time stamp for this session in a StartTime parameter (step 370). The PDSN/FA 130 then sends the LCP Echo Request message to the mobile station 100 (step 375). The mobile station 100 receives the LCP Echo Request message and, in response, returns the LCP Echo Response message to the PDSN/FA 130. The PDSN/FA 130 receives the LCP Echo Response message (step 380) and gets and stores the current time stamp for this session in a StopTime parameter (step 385).

**[0039]** The PDSN/FA 130 then computes the Wireless Access Latency using the StartTime and StopTime parameters (390). The computed Wireless Access Latency is then assigned to the FMLT (FA to MS Latency Time in milliseconds) accounting parameter in, for example, a UDR of the PDSN 130. The PDSN 130 can then send the FMLT to the AAA server 135 in a RADIUS Accounting Request message step 395). Once the message has been received, the AAA server 135 returns a Radium Accounting Response message to the PDSN 130 (step 400).

**[0040]** The latency measurements described herein, including the wireless access latency and the internet access latency measurements provide a way to measure and record user experience in conducting a data call. These measurements can also be used as a way to measure the effect on user experience, if any, of new equipment that is introduced into the network. Using control plane messages to measure latency provides an advantage in that the customer's data usage is not affected. Thus, no billing system changes in the service provider network are needed to incorporate this functionality. Further, by determining latency and therefore quality of service at a per user, per session level, the present invention makes it possible for a service provider to bill a user for providing a particular promised quality of service.

**[0041]** By measuring latency at a per user per session level, a service provider can determine and improve the quality of its wireless infrastructure, and can further improve the

user's experience. Certain technologies, such as Voice over IP (VoIP), real time streaming, etc., that are particularly susceptible to latency will benefit from the ability to measure latency at the user session level.

**[0042]** Although the present invention has been fully described by way of examples and with reference to the accompanying drawings, it is to be understood that various changes and modifications will be apparent to those skilled in the art without departing from the spirit and scope of the invention. Therefore, unless such changes and modifications depart from the scope of the present invention, they should be construed as being included therein.